Liverpool Lung Project lung cancer risk stratification model: calibration and prospective validation

J K Field[1,3], D Vulkan[2,3], M P A Davies[1], S W Duffy[2], R Gabe[2]

[1]The University of Liverpool, Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, The William Duncan Building, 6 West Derby Street, Liverpool, L7 8TX UK.

[2]Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

[3] joint first authors

Correspondence to

Professor John K Field, The University of Liverpool, Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine,The William Duncan Building, 6 West Derby Street, Liverpool, L7 8TX, UK; j.k.field@liverpool.ac.uk

ABSTRACT

BACKGROUND: Early detection of lung cancer saves lives, as demonstrated by the two largest published low-dose CT screening trials. Optimal implementation depends on our ability to identify those most at risk.

METHODS: Version 2 of the Liverpool Lung Project risk score (LLPv2) was developed from case-control data in Liverpool and further adapted when applied for selection of subjects for the UK Lung Screening Trial (UKLS). The objective was to produce version 3 (LLPv3) of the model, by calibration to national figures for 2017. We validated both LLPv2 and LLPv3 using questionnaire data from 75,958 individuals, followed up for lung cancer over 5 years. We validated both discrimination, using ROC analysis, and absolute incidence, by comparing deciles of predicted incidence with observed incidence. We calculated proportionate difference as the percentage excess or deficit of observed cancers compared to those predicted. We also carried out Hosmer-Lemeshow tests.

RESULTS: There were 599 lung cancers diagnosed over 5 years. The discrimination of both LLPv2 and LLPv3 were significant with an area under the ROC curve (AUC) of 0.81 (95% CI 0.79 - 0.82). However, LLPv2 overestimated absolute risk in the population. The proportionate difference was -58.3% (95% CI -61.6% - -54.8%), i.e. the actual number of cancers was only 42% of the number predicted.

In LLPv3, calibrated to national 2017 figures, the proportionate difference was -22.0% (95% CI -28.1% - -15.5%).

CONCLUSIONS: Whilst LLPv2 and LLPv3 have the same discriminatory power, LLPv3 improves the absolute lung cancer risk prediction and should be considered for use in further UK implementation studies.

Key Messages

What is the key question?

The Liverpool Lung Project risk model (LLPv2) has been developed to identify those at high risk of lung cancer for intervention (e.g. by low dose CT in the UKLS trial), here we assess the discriminatory and predictive power of the risk model and propose an updated version (LLPv3) more suitable for a contemporary country-wide population.

What is the bottom line?

The LLPv2 risk model was validated with a significant discriminatory power, but overestimates risk compared to the improved LLPv3.

Why read on?

We demonstrate an improved risk prediction model for lung cancer, by relatively simple *a priori* calibration to updated cancer incidence data.

Number of words 3,940

INTRODUCTION

Evidence from randomised trials estimates a significant 20% reduction in lung cancer mortality with invitation to low dose CT screening [1 2]. In the US, lung cancer screening has been recommended by the United States Preventive Services Task Force (USPSTF) and in other countries there are a growing number of implementation studies aimed at addressing outstanding pragmatic issues of wider roll out for national programmes. An issue of fundamental importance is identification of a population at sufficient risk of disease to optimise detection in those who would benefit from screening for lung cancer, minimise unnecessary investigations for those at low risk and provide a cost-effective programme. While there is ample evidence on the discriminatory power of various risk prediction tools, there is a need to validate these tools on prospectively collected data to estimate their predictive power, as recommended by Toll et al [3] and as performed by Ten Haaf et al [4].

The Liverpool Lung Project risk model was developed on a case-control study in Liverpool and provides a percentage risk score for an individual over the next 5 years, based on questionnaire data [5]. It has been validated in three international data sets and had an AUC of 0.82 in the UK Cohort [6]. It was later amended (LLPv2) due to observations in the UK Lung Screening (UKLS) Trial [7 8] in which it was used to select subjects for low-dose CT screening. In this paper we report on further amendments to the model (LLPv3) calibrated to the whole of England and to more recently available lung cancer incidence data, with prospective validation in a large cohort covering two regions of England.

METHODS

Development and calibration

The LLP risk model was derived from data relating to 579 individuals with lung cancer and 1,157 controls, recruited between 1998 and 2005 in an area of North West England. Variables included in the model were age, sex, cigarette smoking history, prior diagnosis of pneumonia, occupational exposure to asbestos, prior diagnosis of malignant tumour, and family history of lung cancer. The model development process and internal validation have been published elsewhere [5].

A second version of the model (LLPv2) was subsequently developed, following observations in the UKLS that resulted in changes to two variables: previous diagnoses of other lung diseases were considered alongside pneumonia, and cigar and pipe smoking were included alongside cigarettes [8]. The change was prompted by the observation in UKLS that those included on the basis of duration of use of other smoking materials had at least the prevalence of lung cancer as those included with the same duration of cigarette smoking [7]. The numerical parameters of the model remained unchanged. Model formulation is given in Supplementary Material 1. The LLP risk models only use smoking duration, as in common with others [9 10 11], we found this to be the strongest predictor of the various smoking metrics, and other aspects such as amount smoked, were not significant when adjusting for duration [4].

The age-standardised incidence ratio for lung cancer in Liverpool compared to England as a whole for the years 2012-16 was 1.86 [12]. The age-standardised rates of newly diagnosed cases of lung cancer for England were 101.4 per 100,000 of population for males in 2006 and 57.9 for females; the corresponding rates in 2017 (the most recent year for which these data were available) were 86.8 and 67.0 respectively [13]. In order to produce an updated version of the model (LLPv3) for current use in the UK, the age-related risk factors in the model were each adjusted to approximate the effect of reducing the calculated risk score by a factor of 1.86 x 101.4 / 86.8 for males and 1.86 x 57.9 / 67.0 for females, in the absence of any other risk factors (see Supplementary Material 1).

Validation

Participants for the UKLS were recruited from people aged 50-75 (at the time of initial approach) living in the areas covered by six Primary Care Trusts in two regions of the UK (North West England and East of England), by means of a questionnaire sent to just under a quarter of a million people. Recruitment methods and participant rates for UKLS have been reported elsewhere [8]. Ethical approval for the study was given by Liverpool Central Research Ethics Committee in December 2010 (reference number 10/H1005/74).

Of the 75,958 people who expressed an interest in participating in the study (by completing and returning the questionnaire between September 2011 and March 2013), 8,729 were assessed to have a high-risk score (estimated 5% risk or higher over five years) for lung cancer. After exclusions, 4,061 gave informed consent for low dose CT randomisation and biosample provision. Of these, 2,028 were randomised to the CT screening arm (intervention) and 2,027 to the non-screening arm (control); the remaining 6 were not randomised.

The questionnaire sent out as part of the initial approach for recruitment to the trial used a standardised format (see Supplementary Material 2) and covered personal history of non-malignant lung diseases, previous malignancy, exposure to asbestos, family history of lung and other cancers, and smoking history. For all those who had ever smoked regularly, questions covered the type of smoking material, quantity, and the age of starting and (if applicable) stopping smoking. For validation of the amended model (LLPv3), we used prospective data for the 75,958 people who returned the initial questionnaire. Coding and interpretation details are given in Supplementary Material 3.

Cancer and death registration data between 2011 and 2018 were obtained from Public Health England (PHE) and linked through NHS number by a third party to the research cohort (PHE Office of Data Release study 1819_074). This provided the minimum of five years of follow up required for testing the LLP model. Only diagnoses within the first five years were taken into account.

A risk score representing cancer risk over five years was calculated for each individual from the questionnaire data using both LLPv2 and LLPv3. The input variables were derived according to the algorithms shown in the online supplementary material.

A receiver operating characteristic (ROC) curve was plotted to assess the ability of the model to successfully distinguish between those persons who went on to develop lung cancer and those who did not. The ROC curve is a plot of the true positive rate (or sensitivity) against the false positive rate (1 – specificity) for different risk scores. The area under the curve (AUC) was calculated, along with the corresponding 95% confidence interval.

Individuals allocated to the intervention arm of the UKLS trial, nested within the validation cohort, may have potentially received additional diagnostic procedures in the form of low-dose CT scans, compared to the rest of the study population. In order to investigate the possible impact of this, sensitivity analysis was carried out, excluding this subgroup.

In order to assess the calibration of the predictive model, we ranked the LLPv3 risk scores, assigned decile groups, and carried out a linear regression of the log-odds ($f(x) = \log_e \frac{x}{1-x}$) of the rate of observed cancers in each group against the log-odds of the mean of the risk scores in the group. The proportionate difference (the number of cancers observed in excess or deficit of those that would have been expected based on the risk scores, divided by that expected number) was also calculated. In addition, we carried out Hosmer-Lemeshow tests, with the population classified by predicted risk deciles [14]. However, interpretation of these has to be qualified by the test's sensitivity to categorisation, sample size and ties [15] [16] [17].

Statistical analyses were performed using STATA, version 15.1 (StataCorp, College Station, Texas).

RESULTS

Table 1 shows the basic demographic and risk profile of the study population, and of the subgroup developing lung cancer. Overall, almost half the population were male, 65% were aged 60 years or over, 18% were in the most deprived IMD quintile, and more than half the population had smoked at some time in their lives. As expected, the lung cancer cases were older, had higher rates of smoking and were more deprived than the overall population.

Lung cancers were diagnosed in a total of 599 people over the period of follow-up (five years from the questionnaire completion date), i.e. 0.8% of the study population. Table 2 shows the number of cancers that were potentially detectable, i.e. occurring in subjects who qualify for surveillance and those occurring in subjects who do not qualify using the amended LLP model (assuming 100% sensitivity of CT screening) for selected referral thresholds. The table also shows the corresponding sensitivity and specificity of exceeding the threshold. Results are given for males and females separately, and overall. Under LLPv2, a threshold of 3.7% would have yielded 10,864 subjects qualifying. The equivalent threshold under LLPv3 would be 2.0%. The numbers needed to screen (with the risk model) per cancer potentially detectable, fall from 62 with a 0.5% LLPv3 threshold to 21 with a 4.5% threshold. The positive likelihood ratios increased from 2.0 at the LLPv3 0.5% threshold to 6.3 at the 4.5% threshold. The corresponding ranges for males only were 1.9-5.4 and for females 2.3- 7.8.

The ROC curve for LLPv3 is given in Figure 1. The curves for LLPv2 and LLPv3 are almost identical, as would be expected, since the coefficients pertaining to the risk factors are the same, and the ranking of most individuals is unchanged from version 2 to version 3. As a result, the discrimination is very similar. The AUC in each case was 0.81 (95% CI 0.79 - 0.82). After excluding individuals allocated to the UKLS trial intervention arm, the AUC and 95% confidence interval were also unchanged.

The numbers of cancers detected by LLP decile groups and the expected numbers of cancers in each group are given in Table 3, for both LLPv2 and LLPv3. Corresponding figures separately for males and females are given in Supplementary Material 4. The increasing actual risk with predicted risk can be seen from the increasing percentage of cancers in successive risk deciles. The 10% of subjects in the second decile of predicted risk by LLPv3 contribute 2% (12/599) of the cancers whereas the 10% in the 6th decile contribute 4% (24/599) and the 10% in the highest decile contribute 45% (269/599).

Figure 2 shows the log odds of the observed rate of cancers (with 95% confidence intervals) plotted against the log odds of the mean risk score for both LLPv2 and LLPv3.

For LLPv2, the slope of the calibration line was 0.88 (95% CI 0.74 – 1.02), and the intercept at the origin (corresponding to a risk score of 50%) was -1.31 (95% CI -1.96 – -0.66). The orange line represents perfect agreement between observed and expected cancers, indicating that the LLPv2 model overestimated the risk. The proportionate difference was (599-1,436)/1,436 or -58.3% (95% CI -61.6% - -54.8%).

For LLPv3, the slope was largely unchanged at 0.90 (95% CI 0.75 – 1.05), the intercept at the origin was -0.66 (95% CI -1.46 – 0.15), and the proximity to the orange line indicates reduced discrepancy between observed and expected cancers. There was still a tendency to overestimate the number of cancers, especially at the highest risk levels, but the proportionate difference was considerably lower, at -22.0%.

Figure 3 shows the corresponding graphs for LLPv3 for males and females separately. The slope for males was 0.89 (95% CI 0.79 - 0.99) and the intercept was -0.74 (95% CI -1.28 - -0.21). For females, there was stronger agreement between predicted and observed numbers, with a slope of 0.93 (95% CI 0.72 – 1.14) and an intercept of 0.47 (95% CI -1.6 - +0.65).

The corresponding Hosmer-Lemeshow tests yielded a chi-squared of 530.06 ($p<0.001$) for LLPv2 overall, with 437.78 ($p<0.001$) for males and 110.18 ($p<0.001$) for females. For LLPv3, a considerably improved fit was shown, with the Hosmer-Lemeshow statistics reduced by a factor of 10, with chi-squared statistics of 50.87 ($p<0.001$) overall, 43.82 ($p<0.001$) for males and 15.19 ($p=0.06$) for females. In the figures for LLPv3, overall and males only, 65% and 73% respectively of the chi-squared figures were contributed by the highest decile. This can also be seen in Table 3 and Supplementary Material 4.

DISCUSSION

We present an updated version of the LLP model (LLPv3) designed to assess risk of lung cancer in the next five years for the contemporaneous English population. Its primary intended use is to identify people who might benefit from lung cancer screening. We recalibrated the LLPv2 model used in the UKLS trial of lung cancer screening taking account of changes to the age and sex specific lung cancer risk profile of England and the difference in risk between the Liverpool area and England as a whole. We prospectively validated the LLP model by using baseline risk factors and follow-up of lung cancer diagnoses in a large cohort (>75,000) of individuals expressing interest in the UKLS study. This demonstrated excellent discrimination based on an AUC measure of 0.81 (95%CI 0.79 - 0.82) for both LLPv2 and the updated LLPv3. The newer model also provided a very good estimation of future risk through comparison of observed and expected lung cancer outcomes.

However, without the calibration to the current England population in terms of age and sex specific lung cancer risk, there would have been considerable overestimation of risk. This is consistent with the findings of Katki et al [18], who concluded that previous versions of the LLP model overestimated future incidence.

Risk-based lung cancer screening strategies prevent significantly more lung cancer deaths than the current the US Preventive Services Task Force (USPSTF) recommendations on lung cancer screening [19], which is currently the only internationally approved government stratification system and is also supported by the USA medical care agency (Medicare) [20-23]. Two lung cancer risk prediction models have been previously used in lung cancer CT screening clinical trials, the PanCan risk model (precursor to the $PLCO_{M2012}$ model), in the PanCan cohort CT trial [24] and the LLPv2 in the UKLS trial [8]. The PanCan model did show good overall discrimination, however, it underestimated the observed risk of lung cancer by 30%. The authors tested both the PanCan and the $PLCO_{m2012}$ in the PLCO trial “ever smokers” and found small differences in the overall prediction,

discrimination and calibration [24]. A number of other lung cancer risk models have been developed and although some have been tested in previously recruited NLST subjects and a subgroup of the PLCO population, none have been used as the basis for eligibility in the clinical trial environment [4].

The UKLS was a pilot CT screening trial and a pragmatic decision was made to select individuals with a 5% predicted risk of developing lung cancer over a 5-year period, in order to ensure the trial had a large number of lung cancers. This objective was successfully achieved, as 2.2% of subjects with lung cancer were identified in the baseline screen, which was significantly higher than either the NLST [1] or NELSON [2] trials. It was acknowledged that the risk cut off might be lower in a large national screening programme [25].

The discriminatory power of the LLPv2 was significant with an AUC of 0.81, however the model was found to overestimate the absolute risk approximately two fold. The LLP risk model was based on a very high-risk population recruited between 1997-2005 in the NW of England and the incidence rates for lung cancer included in the model were based on 2007 data. On calibrating the age and gender specific intercepts to the current English data [12 13], a recalibrated LLP risk model is now provided here (LLPv3 risk model). As the calibration was independent of the data available from the UKLS study, we were able to use that to validate the model. The recalibrated LLPv3 model demonstrated a significantly improved absolute prediction with the expected rates of cancer in the population, as demonstrated by the correlation between the risk score and the observed cancer rates.

While the prediction as measured by the Hosmer-Lemeshow test is not perfect, we note the reservations about this test specified above. We also note that the LLPv3 clearly predicts better for women than for men, and that the only substantial departure of observed from predicted numbers is in one extreme category, the highest decile of risk, and in males only.

We acknowledge there are a number of limitations, however, many are shared by similar risk model studies. Firstly, the study population, although large at 75,958, represents a self-selecting group from 247,354 individuals (30.7%) approached for the UKLS trial. Potentially, there is over-representation of lower risk, better educated, “worried well” individuals (often associated with higher socio-economic groups) [26]. On examining the UKLS IMD quintiles, the two least deprived (51.2%) outnumber the two most deprived (31.2%); however, it should be acknowledged that the LLP risk score is predictive across all IMD quintiles and a similar IMD profile would be expected on the introduction of national screening. There is a recognition that more should be done to target the “hard to reach” individuals in lower socio-economic groups, who have higher rates of lung cancer (as confirmed here: most deprived quintile 15.4 per 1000 in 5 years, least deprived 5.1).

Secondly, the UKLS study population represents a limited geographical sub-population of the UK (both in the North West and the East of England), and one of the areas overlaps considerably with the area supplying the original estimation data set. However, no individual appears in both estimation and validation sets, and the validation set appears to be broadly representative as demonstrated by the fit achieved for risk calibrated to data for England as a whole. This highlights that basing a risk score on a defined region (i.e. Liverpool) one may overestimate lung cancer risk if that region has a higher incidence than the rest of the country. When extrapolating risk models from smaller studies in defined regions we therefore recommend that this is taken into account.

In terms of discrimination, previous retrospective validation exercises on LLPv2 have shown a range of discriminatory capacities, with AUCs ranging from 0.67 to 0.82 [6]. However, the lower figure pertains to a cohort undergoing surgery, which would not necessarily be representative of lung cancer cases as a whole.

Primary care data is usually not sufficient to supply the risk factors in this model, or in most other lung cancer risk models [18] [4]. The paradigm for selection of individuals for additional lung cancer surveillance or prevention is to use a crude risk criterion such as ever smoking to select candidates from primary care, then carry out a formal risk assessment as part of a lung health check on those candidates using a tool such as one of the LLP or PLCO models. This has been used in the Lung Screen Uptake Trial [27], the Yorkshire Lung Screening Trial [28], and the Manchester [29] and Liverpool [30] demonstration projects.

We note here that the LLPv2 and LLPv3 are essentially aimed at risk of lung cancer in the next five years. They are not diagnostic tests for lung cancer. As such the positive likelihood ratios are low in comparison with some diagnostic tests. The clinical utility of our preferred risk tool, LLPv3 will depend on what intervention might be triggered by a given risk level. The most obvious intervention would be surveillance with low dose CT, but this need not be the only response to lung cancer risk. In this paper, therefore, we have concentrated on accuracy of the prediction, and have not introduced further dimensions such as clinical or public preference. Detailed clinical implications of different risk thresholds using LLPv3 will be the subject of a future paper. This will likely include Decision Curve Analysis and possibly other tools such as Mean Risk Stratification and Incremental Net Benefit [31] [32] [33].

The LLPv2 and $PLCO_{m2012}$ have been utilised and initial data published in three UK lung cancer early detection / CT screening implementation projects, the Liverpool Healthy Lung Programme (LLPv2) [30], the Lung Screen Uptake Trial [27], and the West London Lung Cancer Screening pilot (WLLCS) [34]. All three of these projects successfully selecting high risk patients and identified early stage lung cancers. The Yorkshire Lung Screening trial [28] has also utilised the LLP_{v2} and $PLCO_{m2012}$ risk models, however the results have not as yet been published.

The WLLCS pilot has published its base line data on both the LLP_{v2} and $PLCO_{m2012}$ risk models. The authors reported that 1,159 participants were eligible for a CT scan, of which 451/1,159 (38.9 %) had a LLP_{v2} ≥2.0 % ; only 71/1,159 (6.1 %) had a $PLCO_{M2012}$ ≥1.5 %, lung cancer was detected in 29/1,145 (2.5 %) participants scanned 5/29 participants with lung cancer did not meet a $PLCO_{M2012}$ threshold of ≥1.51 %; however, all had a LLP_{v2} ≥2.0 %.

The Manchester Health Lung Check (MLHC) selected patients for their study with the $PLCO_{m2012}$ risk model. They have recently analysed the performance of both the LLPv2 and $PLCO_{m2012}$ models in their data set [35], utilising ≥2.5% LLP_{v2} risk model and >1.51 % Risk calculated from the $PLCO_{m2012}$ model. In the MLHC 1430 dataset, the authors calculated that 93.5% (58 of the 62 identified) of the lung cancer cases would have been identified by the LLP_{v2} ≥2.5% risk; with a further 272 participants who would have been eligible with LLP_{v2} but ineligible with $PLCO_{m2012}$, however, the outcome of these individuals was unknown.

An international comprehensive analysis of seven lung cancer risk models, (including LLPv1) has been undertaken utilising data from the PLCO and NLST data sets [4]. However a number of the LLP_{v1} parameters were not available; asbestos exposure, nor history of pneumonia, nor the age of diagnosis of the first degree relatives' family history of lung cancer in the PLCO and NLST datasets. Ten Haaf and co-authors reported that all the seven risk models outperformed the NLST trial eligibility criteria over a wide range of risk thresholds in decision curve analysis, with a higher sensitivity for all models and a slightly higher specificity for a number of the models. The $PLCO_{m2012}$ outperformed all of the models in their analysis and both the LLPv1 and the authors' simplified LLP risk model did not perform as well, overestimating the risk. However, the validation exercise used the LLP v1 model and three of the risk parameters were not available in their test sets.

Given that a likely application for lung cancer risk models such as LLPv3 is determination of eligibility for low dose CT screening for lung cancer, it is worth remarking that there is a trade-off between potential increases in cost-effectiveness with higher

risk thresholds and likely comorbidities, which may either impair eligibility for potentially curative surgery in those detected with lung cancer, or otherwise shorten expected life and thus diminish the potential benefits of screening. See Rivera et al, for a thorough exegesis of this issue [36].

This current analysis provides evidence that the LLPv2 and LLPv3 have the same discriminatory power, but the LLPv3 is now recommended for future UK lung cancer CT screening programmes, as it better estimates the absolute lung cancer risk. The NHS England 'Targeted Screening for Lung Cancer with Low Radiation Dose Computed Tomography' protocol [37] has included the use of the $PLCO_{m2012}$ (at a 1.5% risk cut off) and the LLPv2 (at 2.5% risk cut off). It is now recommended that NHS England utilise the LLPv3, which is calibrated to sex and age specific incidence for the whole of England in 2017; a LLPv2 cut off of 2.5% equates to a risk cut off of 1.33% over 5 years for LLPv3, so it is in line with the $PLCO_{m2012}$ of >1.5% over 6 years. To ensure a substantially enhanced risk group, it might be prudent to use a 2.5% threshold using LLPv3 (see Table 2).

In conclusion, discrimination of LLPv2 and LLPv3 was excellent. LLPv3, which was calibrated to contemporary, English incidence, achieved more accurate prediction of absolute incidence, and would be more effective in selecting a high-risk group for surveillance in England today.

Acknowledgements

The authors wish to thank all the individuals who participated in the Pilot UKLS trial. Furthermore, we thank all the individuals involved in the UKLS data management, especially Mrs. Beverley Green and Mr Nial Hodge.

The authors wish to thank Dr Tariq Malik and his colleagues at Public Health England (PHE) for their assistance in providing an update of the UKLS dataset.

The role of Public Health England. This project involves the data derived from patient-level information collected by the NHS, as part of the care and support of cancer patients. The data is collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England Access to the data was facilitated by the PHE Office for Data Release.

Author Contribution:

JK Field - Trial conception and design, data interpretation, manuscript review

SW Duffy - Trial design, statistical analysis, data interpretation, manuscript review

D Vulkan – Risk modelling, statistical analysis, data interpretation, manuscript review

MPA Davies - data interpretation, manuscript review

R Gabe - data interpretation, manuscript review

Funding

Initial development of the LLP was funded by the Roy Castle Lung Cancer Foundation. The UKLS is funded by the Health Technology Assessment programme of the National Institute for Health Research. Michael Davies is a Roy Castle Lung Cancer Foundation Senior Research Fellow. Daniel Vulkan's and Stephen Duffy's contribution to this research was funded by the National Institute for Health Research (NIHR) Policy Research Programme, conducted through the Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis, PR-PRU-1217-21601. The funding source had no role in the design of our analyses, its interpretation, or the decision to submit the manuscript for publication

The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Competing Interests

JK Field: Speaker's Bureau: AstraZeneca; Advisory Board: Epigenomics; NUCLEIX Ltd. AstraZeneca, iDNA; Grant Support: Janssen Research & Development, LLC. No competing interests from all other co-authors.

Exclusive licence

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in Thorax editions and any other BMJPGL products to exploit all subsidiary rights, as set out in our licence.

Ethics Statement

Ethical approval for the study was given by Liverpool Central Research Ethics Committee in December 2010 (reference number 10/H1005/74).

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England journal of medicine* 2011;365(5):395-409. doi: 10.1056/NEJMoa1102873
2. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *The New England journal of medicine* 2020;382(6):503-13. doi: 10.1056/NEJMoa1911793 [published Online First: 2020/01/30]
3. Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61(11):1085-94. doi: 10.1016/j.jclinepi.2008.04.008 [published Online First: 2009/02/12]
4. Ten Haaf K, Jeon J, Tammemagi MC, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. *PLoS Med* 2017;14(4):e1002277. doi: 10.1371/journal.pmed.1002277
5. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* 2008;98(2):270-6. doi: 6604158 [pii] 10.1038/sj.bjc.6604158 [published Online First: 2007/12/19]
6. Raji OY, Duffy SW, Agbaje OF, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. *Ann Intern Med* 2012;157(4):242-50. doi: 10.7326/0003-4819-157-4-201208210-00004 1351360 [pii] [published Online First: 2012/08/23]
7. Field JK, Duffy SW, Baldwin DR, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016;71(2):161-70. doi: 10.1136/thoraxjnl-2015-207140
8. Field JK, Duffy SW, Baldwin DR, et al. The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technol Assess* 2016;20(40):1-146. doi: 10.3310/hta20400
9. International Agency for Research on Cancer (IARC). Tobacco smoking and tobacco smoke 2002 [Available from: http://www.inchem.org/documents/iarc/vol83/01-smoking.html accessed 7/8/2020.
10. Remen T, Pintos J, Abrahamowicz M, et al. Risk of lung cancer in relation to various metrics of smoking history: a case-control study in Montreal. *BMC Cancer* 2018;18(1):1275. doi: 10.1186/s12885-018-5144-5 [published Online First: 2018/12/21]
11. Haws L, Jr., Haws BT. Aerodigestive cancers: lung cancer. *FP Essent* 2014;424:32-47. [published Online First: 2014/09/10]
12. Public Health England. Public Health Profiles 2020 [Incidence of lung cancer, standardised incidence ratio (Liverpool CCG)]. Available from: https://fingertips.phe.org.uk/local-health#page/9/gid/1938133184/pat/46/par/E39000026/ati/154/are/E38000101/iid/93237/age/1/sex/4 accessed 3 April 2020.
13. Office for National Statistics. Cancer registration statistics 2019 [Table 8: Directly age-standardised rates per 100,000 population of newly diagnosed cases of cancer (third digit): site and sex, England 1995 to 2017]. Available from:

https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland accessed 3 April 2020.
14. Hosmer DW, Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. New York: Wiley 2013.
15. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med* 2007;35(9):2052-6. doi: 10.1097/01.CCM.0000275267.64078.B0 [published Online First: 2007/06/15]
16. Kuss O. Global goodness-of-fit tests in logistic regression with sparse data. *Stat Med* 2002;21(24):3789-801. doi: 10.1002/sim.1421 [published Online First: 2002/12/17]
17. Bertolini G, D'Amico R, Nardi D, et al. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. *J Epidemiol Biostat* 2000;5(4):251-3. [published Online First: 2000/10/31]
18. Katki HA, Kovalchik SA, Petito LC, et al. Implications of Nine Risk Prediction Models for Selecting Ever-Smokers for Computed Tomography Lung Cancer Screening. *Ann Intern Med* 2018;169(1):10-19. doi: 10.7326/M17-2701 [published Online First: 2018/05/26]
19. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med* 2014;160(5):311-20. doi: 10.7326/M13-2316
20. Moyer VA, Force USPST. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160(5):330-8. doi: 10.7326/M13-2771 [published Online First: 2014/01/01]
21. Centers for Medicare & Medicaid Services. Decision Memo for Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439N): https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274.
22. Field JK, Aberle DR, Altorki N, et al. The International Association Study Lung Cancer (IASLC) Strategic Screening Advisory Committee (SSAC) Response to the USPSTF Recommendations. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2014;9(2):141-3. doi: 10.1097/JTO.0000000000000060
23. Ten Haaf K, Bastani M, Cao P, et al. A comparative modeling analysis of risk-based lung cancer screening strategies. *J Natl Cancer Inst* 2019 doi: 10.1093/jnci/djz164 [published Online First: 2019/10/01]
24. Tammemagi MC, Schmidt H, Martel S, et al. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. *Lancet Oncol* 2017;18(11):1523-31. doi: 10.1016/S1470-2045(17)30597-1 [published Online First: 2017/10/23]
25. Ten Haaf K, De Koning H, Field J. P2.13-025 Selecting the Risk Cut off for the LLP Model. *Journal of Thoracic Oncology* 2017;12(11):S2174. doi: 10.1016/j.jtho.2017.09.1371
26. Khadjesari Z, Newbury-Birch D, Murray E, et al. Online health check for reducing alcohol intake among employees: a feasibility study in six workplaces across England. *PLoS One* 2015;10(3):e0121174. doi: 10.1371/journal.pone.0121174 [published Online First: 2015/03/24]
27. Quaife SL, Ruparel M, Dickson JL, et al. Lung Screen Uptake Trial (LSUT): Randomized Controlled Clinical Trial Testing Targeted Invitation Materials. *Am J Respir Crit Care Med* 2020;201(8):965-75. doi: 10.1164/rccm.201905-0946OC [published Online First: 2019/12/12]
28. Gabe R, Callister M, Crosbie PA, et al. The Yorkshire Lung Screening Trial 2018 [Available from: https://abstracts.ncri.org.uk/abstract/the-yorkshire-lung-screening-trial/.
29. Crosbie PA, Balata H, Evison M, et al. Implementing lung cancer screening: baseline results from a community-based 'Lung Health Check' pilot in deprived areas of Manchester. *Thorax* 2019;74(4):405-09. doi: 10.1136/thoraxjnl-2017-211377 [published Online First: 2018/02/15]
30. Ghimire B, Maroni R, Vulkan D, et al. Evaluation of a health service adopting proactive approach to reduce high risk of lung cancer: The Liverpool Healthy Lung Programme. *Lung Cancer* 2019;134:66-71. doi: 10.1016/j.lungcan.2019.05.026 [published Online First: 2019/07/20]
31. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565-74. doi: 10.1177/0272989X06295361 [published Online First: 2006/11/14]
32. Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Stat Med* 2001;20(11):1563-74. doi: 10.1002/sim.789 [published Online First: 2001/06/08]
33. Katki HA. Quantifying risk stratification provided by diagnostic tests and risk predictions: Comparison to AUC and decision curve analysis. *Stat Med* 2019;38(16):2943-55. doi: 10.1002/sim.8163 [published Online First: 2019/05/01]

34. Bartlett EC, Kemp SV, Ridge CA, et al. Baseline Results of the West London lung cancer screening pilot study - Impact of mobile scanners and dual risk model utilisation. *Lung Cancer* 2020;148:12-19. doi: 10.1016/j.lungcan.2020.07.027 [published Online First: 2020/08/11]
35. Lebrett MB, Balata H, Evison M, et al. Analysis of lung cancer risk model (PLCOM2012 and LLPv2) performance in a community-based lung cancer screening programme. *Thorax* 2020;75(8):661-68. doi: 10.1136/thoraxjnl-2020-214626 [published Online First: 2020/07/08]
36. Rivera MP, Tanner NT, Silvestri GA, et al. Incorporating Coexisting Chronic Illness into Decisions about Patient Selection for Lung Cancer Screening. An Official American Thoracic Society Research Statement. *Am J Respir Crit Care Med* 2018;198(2):e3-e13. doi: 10.1164/rccm.201805-0986ST [published Online First: 2018/07/14]
37. NHS England National Cancer Programme. Targeted Screening for Lung Cancer with Low Radiation Dose Computed Tomography; Standard Protocol prepared for the Targeted Lung Health Checks Programme 2019 [Available from: https://www.england.nhs.uk/wp-content/uploads/2019/02/targeted-lung-health-checks-standard-protocol-v1.pdf accessed 16/04/2019.

Table 1: Baseline factors and lung cancers diagnosed

	All participants (%)		Participants diagnosed with lung cancer within 5 years (%)		Cumulative incidence per 1,000 of lung cancer over 5 years
All participants	75,958		599		7.9
Sex					
Male	37,525	(49.4)	327	(54.6)	8.7
Female	38,433	(50.6)	272	(45.4)	7.1
Age					
50-54	12,123	(16.0)	21	(3.5)	1.7
55-59	14,359	(18.9)	54	(9.0)	3.8
60-64	19,007	(25.0)	148	(24.7)	7.8
65-69	19,602	(25.8)	224	(37.4)	11.4
70-74	9,993	(13.2)	134	(22.4)	13.4
75-79	874	(1.2)	18	(3.0)	20.6
Region					
North West England	32,609	(42.9)	332	(55.4)	10.2
East of England	43,349	(57.1)	267	(44.6)	6.2
IMD quintile					
1 (most deprived)	13,893	(18.3)	214	(35.7)	15.4
2	9,813	(12.9)	92	(15.4)	9.4
3	13,337	(17.6)	102	(17.0)	7.6
4	16,707	(22.0)	77	(12.9)	4.6
5 (least deprived)	22,198	(29.2)	114	(19.0)	5.1
Unknown	10	(-)	-	(-)	-
Smoking duration					
Never	35,535	(46.8)	58	(9.7)	1.6
1-19 years	10,889	(14.3)	29	(4.8)	2.7
20-39 years	19,284	(25.4)	208	(34.7)	10.8
40-59 years	10,204	(13.4)	302	(50.4)	29.6
60 years or more	46	(0.1)	2	(0.3)	43.5
History of pneumonia or other lung conditions					
No	57,558	(75.8)	351	(58.6)	6.1
Yes	18,400	(24.2)	248	(41.4)	13.5
Personal history of cancer					
No	66,855	(88.0)	493	(82.3)	7.4
Yes	9,103	(12.0)	106	(17.7)	11.6
Asbestos exposure					
No	64,796	(85.3)	471	(78.6)	7.3
Yes	11,162	(14.7)	128	(21.4)	11.5
Family history of lung cancer					
None	63,321	(83.4)	453	(75.6)	7.2
Early onset (before age 60)	4,532	(6.0)	77	(12.9)	17.0
Late onset (on or after age 60)	8,105	(10.7)	69	(11.5)	8.5

Table 2 Sensitivity and specificity by LLPv3 five-year risk threshold

LLPv3 risk score referral threshold	Participants with score at least equal to threshold	Cancers potentially detectable	Cancers not detectable	Number needed to screen per cancer potentially detectable	Sensitivity of threshold	Specificity of threshold	LLPv2 equivalent referral threshold*
(a) Males only							
0.5%	18,190	293	34	62	89.6%	51.9%	1.08%
1%	11,694	256	71	46	78.3%	69.3%	2.15%
1.5%	8,683	225	102	39	68.8%	77.3%	3.21%
2%	6,712	194	133	35	59.3%	82.5%	4.25%
2.5%	5,140	168	159	31	51.4%	86.6%	5.31%
3%	4,022	143	184	28	43.7%	89.6%	6.32%
4.5%	2,104	95	232	22	29.1%	94.6%	9.32%
(b) Females only							
0.5%	14,347	228	44	63	83.8%	63.0%	0.79%
1%	8,333	197	75	42	72.4%	78.7%	1.59%
1.5%	5,425	162	110	33	59.6%	86.2%	2.37%
2%	4,152	135	137	31	49.6%	89.5%	3.15%
2.5%	2,992	114	158	26	41.9%	92.5%	3.94%
3%	2,340	99	173	24	36.4%	94.1%	4.73%
4.5%	1,079	57	215	19	21.0%	97.3%	7.00%
(c) All subjects							
0.5%	32,537	521	78	62	87.0%	57.5%	0.90%
1%	20,027	453	146	44	75.6%	75.0%	1.86%
1.5%	14,108	387	212	36	64.6%	81.8%	2.89%
2%	10,864	329	270	33	54.9%	86.0%	3.70%
2.5%	8,132	282	317	29	47.1%	89.6%	4.78%
3%	6,362	242	357	26	40.4%	91.9%	5.60%
4.5%	3,183	152	447	21	25.4%	96.0%	8.65%

* to achieve at least as many referrals

Table 3 Risk score ranges and observed and expected lung cancers by risk decile for LLPv2 and LLPv3

Decile	LLPv2				LLPv3			
	Size of decile group*	**Range of risk score (%)**	**Number of cancers observed (%)**	**Expected number of cancers**	**Size of decile group***	**Range of risk score (%)**	**Number of cancers observed (%)**	**Expected number of cancers**
1 (lowest risk)	7,790	0.06 – 0.17	0 (0)	10	8,457	0.03 - 0.10	0	6
2	7,928	0.18 – 0.27	12 (2)	18	7,541	0.11 - 0.15	12 (2)	10
3	7,180	0.28 – 0.38	12 (2)	23	6,807	0.16 - 0.20	11 (2)	12
4	7,521	0.39 – 0.49	20 (3)	33	8,284	0.21 - 0.28	25 (4)	20
5	7,622	0.50 – 0.69	20 (3)	45	7,042	0.29 - 0.37	16 (3)	23
6	7,562	0.70 – 1.00	20 (3)	64	7,887	0.38 - 0.56	24 (4)	37
7	7,614	1.01 – 1.61	49 (8)	97	7,154	0.57 - 0.82	36 (6)	49
8	7,555	1.62 – 2.63	72 (12)	152	7,747	0.83 - 1.39	76 (13)	84
9	7,687	2.64 – 5.08	131 (22)	283	7,500	1.40 - 2.64	130 (22)	145
10 (highest risk)	7,499	5.09 – 53.71	263 (44)	711	7,539	2.65 - 34.72	269 (45)	382
All	75,958		599	1,436	75,958		599	768

* these are not equal, since individuals were allocated to groups by an algorithm which ensure all those with the same score were assigned to the same group

FIGURE LEGENDS

Figure 1 ROC curve for LLPv3

Figure 2 Observed vs predicted log odds of cancer for LLPv2 and LLPv3

Figure 3 Observed vs predicted log odds of cancer for LLPv3 (a) Males; (b) Females.